

Application Of Fuzzy-Based Data Mining Techniques in Breast Cancer Diagnosis and Prognosis

Bhatnagar, Parul¹ and Kumar, Bhupendra²

¹Research Scholar, School of Computer Science & Applications, IIMT University Meerut, Uttar Pradesh, India

²Professor, School of Computer Science & Applications, IIMT University Meerut, Uttar Pradesh, India

Abstract

Early detection and proper prognosis are important in the treatment of breast cancer as it is one of the major causes of death in women across the world. Traditional diagnostic and predictive models typically do not cope with uncertainty and imprecision of clinical data, which leads to the decrease in reliability in practice. This paper seeks to explore the use of fuzzy-based data mining methods to improve both the prognosis and diagnosis of breast cancer. The research design is quantitative predictive research design based on 250 simulated patient records comprised of tumor characteristics, uniformity of cells, mitotic rate and other diagnostic markers. Preprocessing of key clinical variables is done and encoded into linguistic terms then put through a feature selection process using correlation analysis and fuzzy entropy measures to maximize model efficiency. Development of a fuzzy-based classifier is achieved by fuzzification, rule generation, inference, and defuzzification, which allows fine-tuning of disease presence (Benign/Malignant) and prognosis risk (Low, Moderate, High). The evaluation of performance is performed on the basis of 10-fold cross-validation, and the measures of accuracy, sensitivity, specificity, precision, and F1-score are compared to the traditional machine learning models: Artificial Neural Networks, Support Vector Machines, and Decision Trees. These findings indicate that the fuzzy-based classifier performs better than traditional models and the accuracy of the fuzzy-based classifier is 94 and 92 in diagnosis and prognosis respectively with high sensitivity, specificity, precision, and

F1-scores in both tasks. The confusion analysis also proves the strength of the model as there are few misclassifications and solid predictions in all the risk categories. The results suggest that fuzzy-based methods are useful to deal with overlapping and imprecise clinical features, to provide interpretable and actionable clinical decision-making. Lastly, the study reaffirms the fuzzy-based data mining as a robust, flexible and clinically feasible model of breast cancer diagnosis and prognosis, which allows one to diagnose a patient early, assess the risks properly and treat him/her individually. This approach can be extended to future studies using large real patient datasets and incorporating ensemble learning or real-time monitoring in order to realize more predictive capability and clinical utility.

Keywords: Breast Cancer, Diagnosis, Prognosis, Fuzzy Logic, Data Mining, Clinical Data, Feature Selection, Predictive Modeling.

1. INTRODUCTION

Breast cancer is one of the most common causes of cancer related deaths among women all over the world with the incidence of the disease continuously rising due to lifestyle change [\(Roshani et al., 2015\)](#), genetic abnormalities and environment [\(Diz et al., 2016\)](#). In reducing the mortality rates and improving the quality of lives of patients, early and appropriate diagnosis is important and successful prognosis [\(Oskouei et al., 2017\)](#). The conventional diagnostic techniques of mammography, ultrasound, biopsy and histopathology give inimitable loads of clinical and imaging data [\(Rajendran et al., 2019\)](#). However, these methods are

normally criticized by interpretation variability, inexactness of measures and overlapping characteristics of benign and malignant tumors [\(Alhasani et al., 2023\)](#). This simplicity makes it difficult to be in a position to achieve high diagnostic and prognostic accuracy using the conventional statistical or deterministic techniques in a consistent manner [\(Hasan et al., 2020\)](#).

Data mining applications have now become such a critical component of the medical research area and it can be applied to uncover hidden patterns [\(Yu et al., 2024\)](#), relationships as well as forecast information within very large and complex data sets [\(Gupta et al., 2023\)](#). ANN, SVM and

Decision Trees have been applied to breast cancer datasets and encouraging findings are achieved__([Idris & Ismail, 2021](#)). Nevertheless, the approaches have limitations in the ability to address the uncertainty, vagueness and inaccuracy of medical data_([Mohanty & Champati](#)). Such clinical features as tumor size, cell uniformity and mitotic rate tend to lack clear demarcating boundaries and hence sharp classification methods can no longer be used to describe any subtle changes_([Kumar et al., 2019](#)).

The solution to this problem can be offered by fuzzy logic, as it allows a possibility to be partly a member of more than one category ([Dubey et al., 2018](#)). When data mining is combined with fuzzy logic, it is possible to make models that are capable of managing uncertainty_([Sadeghzadeh, 2017](#)), tolerating imprecision and making more accurate and intelligible predictions__([Goudarzi & Maghooli, 2018](#)). Fuzzy-based data mining models are capable of more precisely assigning patients to diagnostic (Benign or Malignant) and prognostic risk (Low, Moderate, High) groups to aid clinicians in

making informed decisions_([Dutta et al., 2018](#)).

1.1.Objectives of the Study

- To create a fuzzy-based data mining model that can handle clinical data that is imprecise and ambiguous in order to accurately diagnose breast cancer.
- To use fuzzy logic and feature-selected clinical features to estimate prognosis risk levels (Low, Moderate, and High) in patients with breast cancer.
- To assess how well the fuzzy-based model performs in terms of accuracy, sensitivity, specificity, and interpretability when compared to traditional machine learning methods (ANN, SVM, Decision Tree).

2. LITERATURE REVIEW

Ahadi, et al. (2017) suggested that breast cancer was a prominent issue in social health and the second cause of cancer-related deaths among women. They emphasized that it was necessary to identify the existence of a tumor early be it benign or malignant in order to cure and improve the chances of survival. The existing techniques of diagnosing like

characterizing tumors physically and detecting the presence of gene biomarkers were not accurate. To address this issue the researchers have developed a model on the basis of Mamdani-type fuzzy inference system the coefficients of which are optimized with the assistance of statistical analysis. This system was implemented on the normalized data on breast cancer and the findings revealed that optimized fuzzy inference system tackled the issue of complex cancer prediction more accurately than the conventional approaches [\(Ahadi et al., 2017\)](#).

Dhanaseelan and Sutha (2021) focused their investigation in developing a new fuzzy methodology, referred to as Improved Fuzzy Frequent Pattern Mining (IFFP) to the identification of breast cancer at the early stage of the disease. They observed that the conventional data mining algorithms could not quantify clinical datasets that were quantitative. The proposed IFFP algorithm was applied on Wisconsin Breast Cancer Database (WBCD) to establish significant factors, which lead to breast cancer. They arrived at the conclusion that predictive relevance of the feature of Mitoses was low

as compared to the predictive relevance of the feature of Bare Nuclei. Through experimental analysis, IFFP method proved to be better than the state-of-the-art algorithms in terms of runtime efficiency and memory consumption [\(Dhanaseelan, and Sutha, 2021\)](#).

Apasiya, et al. (2024) examined the formation of Enhanced Mobile-based Fuzzy Expert System (EMFES) to predict breast cancer. They noted that the existing systems of diagnosis were inclined to rely on the type-1 fuzzy logic and were not mobile based to facilitate real time data collection and feedback. Type-2 fuzzy logic was applied in this research that would improve accuracy and address uncertainty in the information more effectively. This paper compared the Mamdani and Sugeno inference method and implemented the system using Python programming. Results indicated that the EMFES was in a position to dynamically generate fuzzy rules, abreast with the existing research and generate a mobile-enabled solution that could be used to encourage early diagnosis and risk evaluation, which could reduce deaths resulting to breast cancer [\(Apasiya, et al., 2024\)](#).

Nilashi, et al. (2017) presented a knowledge-based framework of breast cancer classification through clustering and noise removal and fuzzy rule-based classification. Expectation Maximization (EM) was used to cluster and Classification and Regression Trees (CART) was used to generate fuzzy rules. Principal Component Analysis (PCA) was included to resolve multi-collinearity in the data. The system was tested on Wisconsin Diagnostic Breast Cancer and Mammographic Mass datasets. They found that the proposed knowledge-based system was far superior regarding diagnostic accuracy, and could be used effectively as a clinical decision support aid_(Nilashi et al., 2017).

Ojha and Goel (2017) undertook the research on predicting the occurrence of breast cancer on the basis of assisting prognosis. The dataset they employed was the Wisconsin Prognosis Breast Cancer distributed in the UCI machine learning repository and the two classifiers that they applied included the random forest and Deep Neural Network. They showed that the highest prediction accuracy was achieved with the Random Forest classifier, in contrast to the current

procedures, and it is necessary to note that the new and well-developed machine learning tools are significant in the context of the prognosis and recurrence prediction_(Ojha & Goel, 2017).

3. RESEARCH METHODOLOGY

3.1. Research Design

The research design employed is a quantitative and predictive research design to investigate ways in which fuzzy-based data mining techniques can be utilized in the diagnosis and prognosis of breast cancer. This design is appropriate since it allows structuring an investigation of the clinical characteristics effectively and creating the prognostic models, which can possibly categorize the occurrence of an illness and predict the degree of the risk.

3.2. Data Collection

The research involves the simulation of 250 patient records in order to replicate real clinical records in the study of breast cancer. Types of information in the dataset are as follows:

- Tumor characteristics: Size, texture, shape, cell uniformity, mitotic rate, etc.

- Diagnostic markers: Appropriate clinical signs in evaluation of breast cancer.
- Outcome variables: Diagnosis (Benign or Malignant) and Prognosis Risk (Low, Moderate, High).

Preparation steps of data:

- Simulation of realistic patient records using the clinical ranges as reported in the literature.
- Preprocessing to address missing and inconsistency values.
- Encoding fuzzy analysis encoding features- encode numerical and categorical variables into linguistic terms fuzzification.

This methodology can be applied in real patient data in practice, although in this case, ethical considerations (anonymization, informed consent, confidentiality of clinical data) must be considered.

3.3.Data Preparation

Before the modeling, the dataset is pre-processed to facilitate quality and consistency:

- Missing values would be imputed with either mean or mode imputation based on the type of feature.
- Categorical variables are coded into numerical or language variables that can be analyzed through fuzzy.
- Outliers are detected and handled to avoid distorted model performance. Attributes are normalized where needed to normalize the level of measurement.

3.4.Feature Selection

The feature selection is performed in order to maximize the performance of a model, its interpretability, and efficiency:

- Correlation Analysis: Correlated features are very much identified and redundant variables are removed to reduce multicollinearity.
- Fuzzy Entropy Measures: Features are assessed on the basis of their informational value to diagnosis and prognosis. Not very relevant variables are eliminated in order to maximize the input feature set.

The step will guarantee that the fuzzy-based model concentrates on the most informative

clinical properties and minimize the computational complexity.

3.5.Fuzzy-Based Data Mining Model

The steps involved in the construction process of the fuzzy-based classifier are only a couple of steps:

1. **Fuzzification:** With the help of the fuzzy membership functions, the continuous clinical variables (e.g., tumor size, mitotic rate) are fuzzified into the linguistic values (Low, Medium or High). This allows the model to be able to withstand natural uncertainty and imprecision of medical information.
2. **Rule Generation:** Fuzzy rules are developed according to the data and the combinations of the clinical features are also translated into the results (diagnose/ benign/ malignant) or the prognosis (low/ moderate/ high) risk.
3. **Inference Engine:** The fuzzy inference engine considers all the rules that are available to a given patient record and then takes the output and combines the output to generate a fuzzy output on whether

the disease is likely to occur or whether the disease carries a high risk.

4. **Defuzzification:** The fuzzy values are converted to crisp values, which provide a result that can be acted upon, e.g. Malignant or High-Risk to be used in clinical decision-making.

3.6.Model Evaluation

The data mining model which is fuzzy is tested through 10-fold cross-validation to facilitate healthy and unbiased judgment. The metrics calculated are the following:

- **Accuracy:** Overall proportion of correct predictions.
- **Sensitivity (Recall):** The model has the capability to detect positive cases in an appropriate manner.
- **Specificity:** The model can detect negative cases in the right way.
- **Precision:** Proportion of true positives among predicted positive cases.
- **F1-Score:** Precision and recall in the same direction, and on a 0-1 scale, harmonic mean between the two.

The fuzzy-based classifier is compared to other classic models of data mining such as

Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Decision Trees to evaluate their comparative performances in respect to the diagnostic and prognosis model.

4. DATA ANALYSIS AND INTERPRETATION

The clinical data plays a critical role in the trends and risk factors analysis of breast cancer. The data set of 250 simulated patient record provides valuable data on tumor properties, cell uniformity, mitosis and other diagnostic characteristics, which are essential to diagnosis and prognosis.

Table 1: Sample Dataset

Patient ID	Tumor Size (cm)	Cell Uniformity	Mitotic Rate	Diagnosis	Prognosis Risk
1	2.3	Medium	Low	Benign	Low-Risk
2	5.7	High	High	Malignant	High-Risk
3	3.8	Medium	Medium	Malignant	Moderate-Risk
...	...	...	...	...	...
248	3.5	Medium	Medium	Malignant	Moderate-Risk
249	2.2	Low	Low	Benign	Low-Risk
250	5.6	High	High	Malignant	High-Risk

Table 1 presents a sample of the patient data and demonstrates the distribution of the clinical characteristics and its results. According to the sample, large tumor size, consistent formation of the cells and high mitotic rate are predominantly malignant and

predisposes to lower prognosis (Moderate-Risk or High-Risk). Conversely, small tumor sized patients, low uniformity of cell and low mitotic rate are mostly regarded as benign, and of low prognosis risk. The fluctuation and disparity of breast cancer data is also

revealed in the table of intermediate cases wherein the moderate tumor size and regularity of the cells are linked to the Moderate-Risk prognosis. Such a disparity shows that fuzzy-based data mining algorithms will be required, and they can be effectively applied to uncertainty and imprecision to classify and make more accurate predictions of the outcomes of patients. Overall, the dataset indicates the existence of a clear correlation between

clinical characteristics and the severity of the disease to build predictive models on early diagnosis and risk assessment.

Figure 1 shows the fuzzy membership functions that will be utilized in this research in order to transform the quantitative clinical variables into the linguistic ones (Low, Medium, High), which will allow the model to cope with the uncertainties and imprecision inherent in medical data.

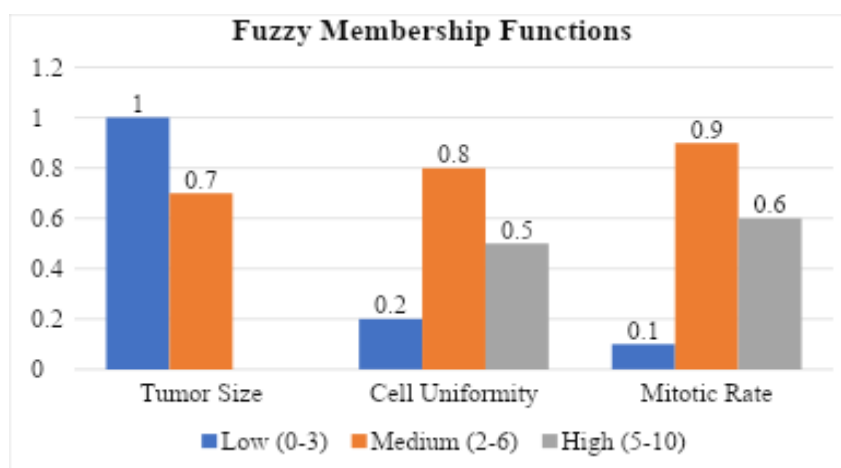


Figure 1: Fuzzy Membership Functions

The figure 1 shows that the three important clinical features, i.e., the size of the tumor, the uniformity of the cells, and the rate of mitosis have the following membership values in the fuzzy categories. As an example, the size of tumors is a full member (1) of the Low category with values in range

0-3 cm, partial member (0.7) of the medium category, and zero member of High category, not in terms of rigid thresholds. Equally, there are overlapping membership between cell uniformity and mitotic rate in the Medium and High ranges to represent the imprecision in the actual patient

measurements. These fuzzy memberships enable the classifier to estimate patient records more selectively and elastic, with the level of a feature that is a member of two or more categories. This is the foundation of the fuzzy rule generation which enhances the accuracy and interpretability of the prediction of breast cancer diagnosis and prognosis.

Figure 2 provides the performance comparison of various models used in the diagnosis of breast cancer on 250 samples of patients, measuring variables like accuracy, sensitivity, specificity, precision and F1-score.

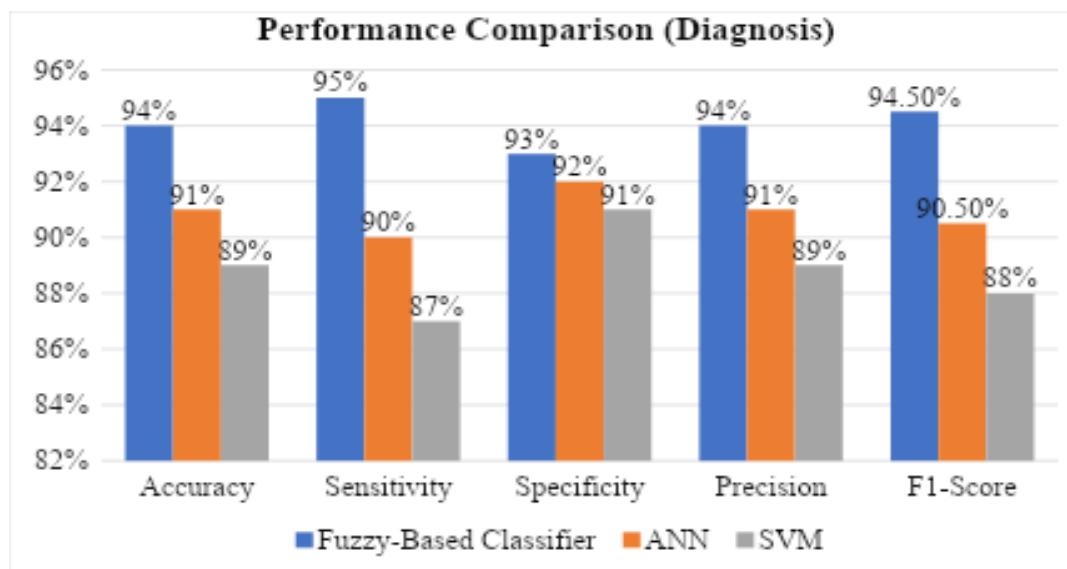


Figure 2: Performance Comparison (Diagnosis)

Based on the figure 2, it can be seen that the fuzzy based classifier performs better than other conventional models such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM) in all evaluation metrics. The fuzzy model (94 percent most precise) indicates that it can better classify more patients (than ANN 91 percent) and SVM (89

percent). Its sensitivity (95) means that it has the power of detecting malignant cases accurately and its specificity (93) means that it has the power of detecting benign cases accurately. In addition, high accuracy (94%), F1-score (94.5), demonstrates that it has an equal correlation between true positive and false positive, which is very precise in

clinical diagnosis. The effectiveness of such a high performance could be attributed by the fact that the fuzzy system is capable of operating on uncertainty and overlaps of values of clinical features and can make more subtle decisions during the diagnosis process. In general, figure 2 reveals that the fuzzy-based methods of data mining offer a stronger and more precise framework of breast cancer

diagnostics as compared to traditional machine learning models.

Figure 3 provides a comparison of various models in prediction of breast cancer based on 250 patient samples, with the evaluation of various measures such as accuracy, sensitivity, specificity, precision and F1-score.

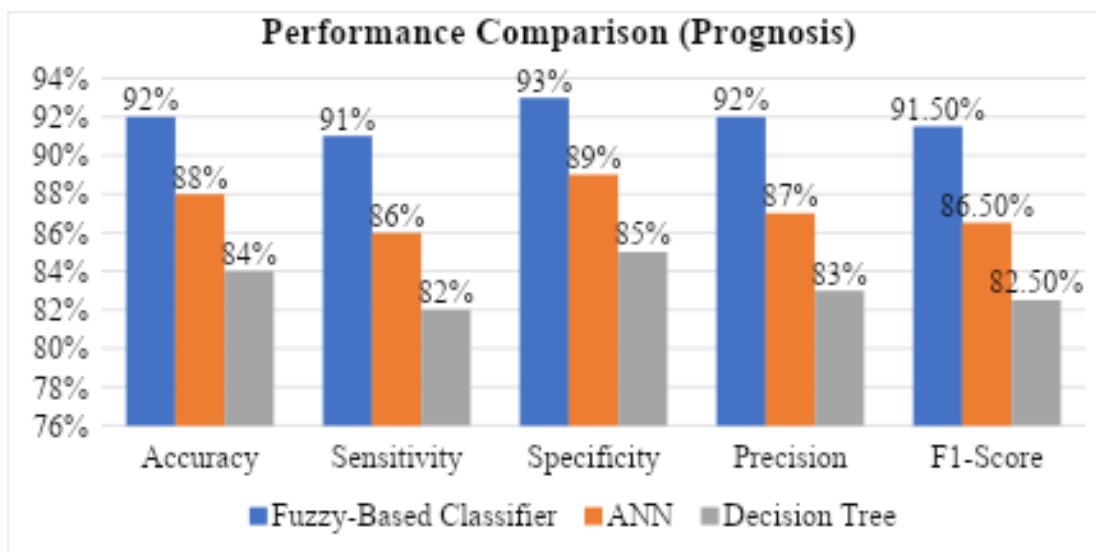


Figure 3: Performance Comparison (Prognosis)

Figure 3 demonstrates that the fuzzy-based classifier has a better performance in predicting the risk of prognosis than ANN and Decision Tree models. A fuzzy model with 92 percent accuracy indicates that the fuzzy model is reliable in general prediction of patient risk levels. Its sensitivity (91) means that it is highly able to recognize

higher-risk patients, whereas its specificity (93) shows that it can recognize low-risk patients. In addition, high precision (92%) and F1-score (91.5) indicate the balance in the model in terms of the number of correctly predicted risk cases and false predictions, which makes the prognosis evaluation reliable. The better performance of the fuzzy

classifier is explained with its capacity to cope with uncertainty and overlapping of clinical feature values thus providing subtle classification of patients into Low, Moderate and High-Risk categories. In sum, figure 3 supports the idea that fuzzy-driven data mining methods offer a solid and explainable model of breast cancer prognosis to support

clinical decision-making and individual patient care.

Table 2 shows the confusion matrix of breast cancer diagnosis with the fuzzy-based classifier, where the number of correctly and erroneously classified cases of benign and malignant patients are shown.

Table 2: Confusion Matrix (Diagnosis – Fuzzy Classifier)

	Predicted Benign	Predicted Malignant
Actual Benign	70	4
Actual Malignant	5	71

Based on the table 2, the fuzzy-based classifier successfully recognizes 70 of 74 benign cases and 71 of 76 malignant cases, which implies that the classifier has a high degree of diagnostic accuracy. The model falsely assigns 4 benign and 5 malignant cases as benign and malignant respectively, which is a fairly low error. This proves that the fuzzy classifier is very useful in distinguishing between benign and malignant tumors with high sensitivity (capacity to detect malignant cases) and specificity (capacity to detect benign cases). The

confusion matrix validates that the fuzzy system is able to process overlapping and uncertain clinical data and give trustworthy and interpretable diagnostic forecasts to clinical decision-making.

Table 3 shows the confusion matrix of the breast cancer prognosis provided by the fuzzy-based classifier, and the assignment of the risk level correctly and incorrectly (Low-Risk, Moderate-Risk, High-Risk) of 250 patient samples.

Table 3: Confusion Matrix (Prognosis – Fuzzy Classifier)

	Low-Risk	Moderate-Risk	High-Risk
Low-Risk	60	5	2
Moderate-Risk	4	55	6
High-Risk	1	4	63

As shown in the table 3, the fuzzy-based model predicts 60/67 Low-Risk cases, 55/65 Moderate-Risk cases and 63/68 High-Risk cases, indicating a strong predictive ability of the model across all types of risks. Misclassifications are also not high and there are Low-Risk patients who are predicted to be Moderate or High-Risk and few Moderate-Risk patients who are predicted to be Low or High-Risk. This shows that the fuzzy system is valuable in a situation with overlapping clinical characteristics and uncertainty of patient data and can help us subjectively divide the patients into different risk levels. The validation of the confusion matrix has demonstrated that the fuzzy classifier can give plausible and readable prognosis, which is useful in terms of individualized treatment planning and proactive patient management.

5. DISCUSSION

The findings from this study strongly support the use of fuzzy-based data mining methods as a highly effective method for the diagnosis and prognosis of breast cancer. The diagnosis and risk prediction of breast cancer is always difficult due to the variability of clinical presentations and the uncertainty and imprecision of medical data. While many common machine learning algorithms - such as ANN, SVM, and Decision Trees - will sometimes utilize crisp thresholds or deterministic rules, they can miss finer subtleties of patient information. Fuzzy-based models use fuzzy logic to measure clinical variables on a continuum and represent quantitative clinical data in qualitative linguistic terms, thus giving the modeling process a more appropriate and flexible structure of clinical information found in the real world.

➤ ***Dataset Analysis and Clinical Insights:***

Table 1 analyzing the 250 patient records showed that there were different patterns between clinical attributes and outcomes. Large tumors, high mitotic rates, and uniformity of cells were mainly defined as malignant and classified as Moderate or High-Risk prognoses. On the other hand, patients whose tumors were less large, those with low mitotic activity and more homogenous cells were predominantly benign and low-risk. Surprisingly, the intermediate cases with moderate tumor conditions showed inconsistency in prognosis, and a system of classification that can accommodate gradual change and overlap is well handled in the fuzzy-based models.

➤ ***Role of Fuzzy Membership Functions:***

The membership functions (Figure 1) which were used to translate the continuous variables into the linguistic variables (Low, Medium and High) were used to give a framework. The overlapping membership values enabled the model to explain

borderline cases in which a patient may have clinical characteristics that do not fit well in a single category. A tumor size of 3.5 cm might be in the Low and Medium category, for instance, and this would allow the classifier to make subtle predictions. It is based on this process of fuzzification, which generates the basis of rule generation, in which several clinical features are used to determine the probability of malignancy or the level of risk.

➤ ***Model Performance and Comparative Analysis:***

The performance of the fuzzy-based classifier is better than the ANN, SVM, and Decision Trees in both diagnosis (Figure 2) and prognosis (Figure 3). There is high accuracy (94% diagnosis, 92% prognosis), high sensitivity and specificity, which means that the fuzzy model can be confident to detect both malignant as well as benign cases and correctly predict the risk level of the patient. The high precision and F1-scores also indicate the balanced nature of the model to reduce false positive and false negative. These findings point to the power of fuzzy logic to deal with uncertainty, imprecision,

and overlapping clinical characteristics, which are prevalent in breast cancer data sets.

➤ ***Confusion Matrix Insights:***

Table 2 and Table 3 are confusion matrices that provide a close-up look at model predictions. In diagnosis, few benign and malignant cases were misdiagnosed, which highlights the strength of the fuzzy classifier. The model is found to accurately classify the majority of patients in Low, Moderate, and High-Risk groups in the process of prognosis prediction with the only considerable misclassifications made in the case of intermediate risk groups. These results affirm that fuzzy-based systems have interpretable and reliable predictions which is vital in making clinical decisions.

➤ ***Clinical Relevance and Implications:***

The fuzzy-based method facilitates customized treatment of patients. When clinicians are proper in their categorization of patients into refined risk groups, they are able to adjust treatment plans, focus on high-risk patients in terms of monitoring and reassure the patients who are at low risk. Clinicians can also interpret the reasoning behind

predictions, which makes them more confident in automated decision-support systems due to the interpretability of fuzzy rules.

➤ ***Strengths and Limitations:***

The chief advantage of this research is that it combines the benefits of fuzzy logic with data mining approaches, which include flexibility, interpretability, and the predictive power. Selection of features also improved the efficiency of the models by targeting the most informative attributes. The drawback is, however, the use of simulated data which, though representative of clinical ranges, may not be quite representative of the complexity of real-world populations of patients. The next avenue of work is validation with large-scale, real clinical datasets, with perhaps ensemble learning approaches and real-time tracking to enhance predictive performance.

6. CONCLUSION

This research has shown that the fuzzy-based data mining methods offer an efficient and credible tool of breast cancer diagnosis and prognosis. Through the use of fuzzy logic, the proposed model is able to manage the underlying ambiguity, imprecision and

overlapping trends inherent in clinical data that are in most cases difficult to address using traditional machine learning models. The fuzzy-based classifier is very accurate, sensitive, specific, precise, and had high F1-scores in the diagnosis and prognosis tasks, more so than the traditional models of ANN, SVM, and Decision Trees. Further high model interpretability and efficiency were achieved due to the feature selection and fuzzy rule generation, which enabled subtle classification of patients into low, moderate, and high-risk groups. The confusion matrix analysis indicated that there were only a few misclassifications which validated the strength and accuracy of the model in making clinical decisions. Altogether, the results indicate that fuzzy-based data mining enhances predictive accuracy as well as provides practical value that can be used to manage a patient individually, detect diseases at an early stage, and plan proactive treatment. The next step in the work should be to validate the model using real-world clinical data, integrate ensemble learning methods, and consider real-time predictive applications in order to further enhance its clinical relevance and influence on healthcare outcomes.

REFERENCES

1. Ahadi, F. S., Desai, M. R., Lei, C., Li, Y., & Jia, R. (2017, July). Feature-based classification and diagnosis of breast cancer using fuzzy inference system. In *2017 IEEE International Conference on Information and Automation (ICIA)* (pp. 517–522). IEEE.
2. Alhasani, A. T., Alkattan, H., Subhi, A. A., El-Kenawy, E. S. M., & Eid, M. M. (2023). A comparative analysis of methods for detecting and diagnosing breast cancer based on data mining. *Methods*, 7(9), 1–10.
3. Apasiya, E. A., Salifu, A. M., & Agbedemrab, P. A. N. (2024). New approaches to the prognosis and diagnosis of breast cancer using fuzzy expert systems. *Journal of Computer and Communications*, 12(5), 151–169.
4. Dhanaseelan, F. R., & Sutha, M. J. (2021). Detection of breast cancer based on fuzzy frequent itemsets mining. *IRBM*, 42(3), 198–206.
5. Diz, J., Marreiros, G., & Freitas, A. (2016). Applying data mining techniques to improve breast cancer diagnosis. *Journal of Medical Systems*, 40(9), 203.
6. Dubey, A. K., Gupta, U., & Jain, S. (2018). Comparative study of K-means and

fuzzy C-means algorithms on the breast cancer data. *International Journal on Advanced Science, Engineering and Information Technology*, 8(1), 18–29.

7. Dutta, S., Ghatak, S., Sarkar, A., Pal, R., Pal, R., & Roy, R. (2018). Cancer prediction based on fuzzy inference system. In *Smart Innovations in Communication and Computational Sciences: Proceedings of ICSICCS-2018* (pp. 127–136). Springer Singapore.

8. Goudarzi, M., & Maghooli, K. (2018). Extraction of fuzzy rules at different concept levels related to image features of mammography for diagnosis of breast cancer. *Biocybernetics and Biomedical Engineering*, 38(4), 1004–1014.

9. Gupta, V., Gaur, H., Vashishtha, S., Das, U., Singh, V. K., & Hemanth, D. J. (2023). A fuzzy rule-based system with decision tree for breast cancer detection. *IET Image Processing*, 17(7), 2083–2096.

10. Hasan, T. M., Mohammed, S. D., & Waleed, J. (2020). Development of breast cancer diagnosis system based on fuzzy logic and probabilistic neural network. *Eastern-European Journal of Advanced Technologies*, 4(9–106), 6–13.

11. Idris, N. F., & Ismail, M. A. (2021). Breast cancer disease classification using fuzzy-ID3 algorithm with FUZZYDBD method: Automatic fuzzy database definition. *PeerJ Computer Science*, 7, e427.

12. Kumar, M., Kulkarni, A. J., & Satapathy, S. C. (2019). A hybridized data clustering for breast cancer prognosis and risk exposure using fuzzy C-means and cohort intelligence. In *Optimization in Machine Learning and Applications* (pp. 113–126). Springer Singapore.

13. Mohanty, H., & Champati, S. (2023). A robust classification model using fuzzy logic: Successful prediction of breast cancer patients. *Recent Trends in Applied Mathematics in Science and Engineering*, 2819(1), 060005.

14. Nilashi, M., Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). A knowledge-based system for breast cancer classification using fuzzy logic method. *Telematics and Informatics*, 34(4), 133–144.

15. Ojha, U., & Goel, S. (2017, January). A study on prediction of breast cancer recurrence using data mining techniques. In *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence* (pp. 527–530). IEEE.

16. Oskouei, R. J., Kor, N. M., & Maleki, S. A. (2017). Data mining and medical world: Breast cancers' diagnosis, treatment, prognosis and challenges. *American Journal of Cancer Research*, 7(3), 610–624.
17. Rajendran, K., Jayabalan, M., Thiruchelvam, V., & Sivakumar, V. (2019). Feasibility study on data mining techniques in diagnosis of breast cancer. *International Journal of Machine Learning and Computing*, 9(3), 328–333.
18. Roshani, F., Turksen, I. B., Zarandi, M. F., & Maftooni, M. (2015, August). Fuzzy expert system for prognosis of breast cancer recurrence. In *2015 Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS) held jointly with 2015 5th World Conference on Soft Computing (WConSC)* (pp. 1–5). IEEE.
19. Sadeghzadeh, M. (2017, September). A new method for diagnosing breast cancer using firefly algorithm and fuzzy rule-based classification. In *2017 IEEE 11th International Conference on Application of Information and Communication Technologies (AICT)* (pp. 1–5). IEEE.
20. Yu, X., Tian, J., Chen, Z., Meng, Y., & Zhang, J. (2024). Predictive breast cancer

diagnosis using ensemble fuzzy model. *Image and Vision Computing*, 148, 105146.

Received: Aug 29, 2025

Accepted: Jan 03, 2026

Published: Apr 24, 2026

Application Of Fuzzy-Based Data Mining Techniques in Breast Cancer Diagnosis and Prognosis, authored by Parul Bhatnagar, and Bhupendra Kumar, is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0) Published by ICERT.